ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

Characterization of Conserved Toxicogenomic Responses in Chemically Exposed Hepatocytes across Species and Platforms

Nehme El-Hachem, Patrick Grossmann, Alexis Blanchet-Cohen, Alain R. Bateman, Nicolas Bouchard, Jacques Archambault, Hugo J.W.L. Aerts, and Benjamin Haibe-Kains

http://dx.doi.org/10.1289/ehp.1409157

Received: 1 September 2014
Accepted: 9 July 2015
Advance Publication: 14 July 2015

Environ Health Perspect DOI: 10.1289/ehp.1409157
Advance Publication: Not Copyedited

Characterization of Conserved Toxicogenomic Responses in Chemically Exposed Hepatocytes across Species and Platforms

Nehme El-Hachem[1,2,$], Patrick Grossmann[$,3,6], Alexis Blanchet-Cohen[4], Alain R. Bateman[5], Nicolas Bouchard[2,8], Jacques Archambault[9], Hugo J.W.L. Aerts[3,6,7,*], and Benjamin Haibe-Kains[10,11,*]

[1]Integrative systems biology, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada; [2]Department of Medicine, University of Montreal, Montréal, Quebec, Canada; [3]Department of Biostatistics & Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; [4]Bioinformatics, Institut de Recherches Cliniques de Montréal, Montreal, Canada; [5]Department of Human Genetics, McGill University, Montreal, Quebec, Canada; [6]Departments of Radiation Oncology Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [7]Department of Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [8]Molecular Biology of Neural Development, Institut de Recherches Cliniques de Montréal, Montreal, Canada; [9]Laboratory of Molecular Virology, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada; [10]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada; [11]Medical Biophysics Department, University of Toronto, Toronto, Ontario, Canada. [$]Co-first authors. [*]Co-last authors.

Address correspondence to Benjamin Haibe-Kains, Princess Margaret Cancer Centre, University Health Network, 101 College Street Toronto, ON, M5G 1L7, Canada. Telephone:

+1 (416) 581-7628. E-mail: bhaibeka@uhnresearch.ca. Hugo J.W.L. Aerts, Department of Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 USA. E-mail: hugo@jimmy.harvard.edu

Running title: Toxicogenomics across species and platforms

Acknowledgments: The authors would like to thank the investigators of the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs) study who have made their invaluable data available to the scientific community.

Funding. NE-H was supported by an IRCM doctoral fellowship. BHK was supported by the Gattuso Slaight Personalized Cancer Medicine Fund at Princess Margaret Cancer Centre. HJWLA was supported by diXa, a part of the EU Seventh Framework Programme, under grant agreement number RI-283775. Authors acknowledge financial support from the National Institute of Health (NIH-USA U01CA190234).

Authors' contributions. NEH and PG participated to the overall design of the study, and were responsible for the collection and normalization of microarray samples, statistical analysis, implementation and documentation of the related R codes, manuscript writing, and the interpretation of the analysis results. HJWLA and BHK designed and supervised the study, and were responsible for the interpretation of the analysis results and manuscript writing. AB-C, ARB, NB and JA assisted and contributed to the interpretation of the results and participated to the manuscript writing. All authors read and approved the final manuscript.

Competing financial interests: The authors declare they have no actual or potential competing financial interests.

ABSTRACT

Background: Genome-wide expression profiling is increasingly being used to identify transcriptional changes induced by drugs and environmental stressors. In this context the TG-GATEs project (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system) generated transcriptional profiles from rat liver samples and human/rat cultured primary hepatocytes exposed to more than 100 different chemicals.

Objectives: To assess the capacity of the cell culture models to recapitulate pathways induced by chemicals *in vivo*, we leveraged the TG-GATEs dataset to compare the early transcriptional responses observed in the liver of rats treated with a large set of chemicals to those of cultured rat and human primary hepatocytes challenged with the same compounds *in vitro*.

Methods: We developed a new pathway-based computational pipeline that efficiently combines gene set enrichment analysis (GSEA) using Reactome pathways and biclustering to identify common modules of pathways that are modulated by several chemicals *in vivo* and *in vitro* across species.

Results: We found that chemicals induce conserved patterns of early transcriptional responses in *in vitro* and *in vivo* settings, and across human and rat. These responses involved pathways of cell survival, inflammation, xenobiotic metabolism, oxidative stress, and apoptosis. Moreover, our results support TGF-beta receptor signalling pathway as a candidate biomarker associated with exposure to environmental toxicants in primary human hepatocytes.

Conclusions: Our integrative analysis of toxicogenomics data provides a comprehensive overview of biochemical perturbations affected by a large panel of chemicals. Furthermore, we show that the early toxicological response occurring in animals is recapitulated in human and rat primary hepatocyte cultures at the molecular level, indicating that these models reproduce

key pathways in response to chemical stress. These findings expand our understanding and interpretation of toxicogenomics data from human hepatocytes exposed to environmental toxicants.

INTRODUCTION

Humans are exposed to a variety of toxic chemicals and have access to a wide array of drugs each of which have the potential to cause short and long-term adverse effects including lethality. From an environmental health perspective, it is important to find a strong connection between toxic substances and human disease susceptibility, therefore elucidating molecular mechanisms of toxicity.

Although animal models are currently the gold standard in evaluating risk and predicting adverse human health effects, they require considerable time and resources, and also raise ethical issues. (Bissell et al. 2001; Greaves et al. 2004; Kola and Landis 2004; Metushi and Uetrecht 2014; Suter et al. 2011) reviewed in (Hebels et al. 2014). For these reasons, several efforts have been made to minimize the use of animals in toxicology (http://www.alttox.org) and to develop robust *in vitro* models predictive of toxicity in human (Abbott 2005). A European initiative, the REACH (Registration, Evaluation, Authorization and Restriction of chemicals) legislation, suggests the use of high-throughput "omics" technologies, such as genome-wide gene expression profiling, to find alternatives to animal testing. The REACH legislation states:

"*The Commission, Member States, industry and other stakeholders should continue to contribute to the promotion of alternative test methods on an international and national level including computer supported methodologies, in vitro methodologies, as appropriate, those based on toxicogenomics, and other relevant methodologies*" (REACH).

Multiple studies used gene expression profiles to characterize toxicogenomic response (Ellinger-Ziegelbauer et al. 2008; Nuwaysir et al. 1999), as reviewed in (Afshari et al. 2011;

Chen et al. 2012). To confront chemical induced cellular stress, the biological system executes a transcriptional control over several signaling pathways (Grinberg et al. 2014; Kier et al. 2004). Because the liver plays a primordial role in detoxification and is a major site of frequent chemical-induced injuries, it was extensively studied in toxicogenomics. Recently, the Japanese government and the pharmaceutical industry joined forces to create and make publicly available the largest toxicogenomic database to date: the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs) (Uehara et al. 2010; Uehara et al. 2011). The TG-GATEs consortium tested approximately 150 chemicals in different models, including primary human and rat hepatocytes as well as rat liver and kidney *in vivo* models (Uehara et al. 2010; Uehara et al. 2011). The experimental design and gene expression profiles were made publicly available through the EBI ArrayExpress database http://www.ebi.ac.uk/arrayexpress/ (Brazma et al. 2003). Different studies used this large toxicogenomic dataset to identify predictive biomarkers of hepatocarcinogenicity (Caiment et al. 2014; Yamada et al. 2012), phospholipidosis (Hirode et al. 2008), and coagulopathy (Hirode 2008). However, despite the availability of these valuable data, it remains unclear whether animal studies could be efficiently replaced by *in vitro* testing to identify key biological pathways induced by hepatotoxic chemicals, one of the main challenges of toxicogenomics.

In this study, we performed a large-scale comparative analysis of the TG-GATEs data from rat liver samples (referred to as RLV) and from cultured rat and human primary hepatocytes (referred to as PRH and PHH) in order to (*i*) identify conserved transcriptional responses induced by chemicals across species and between *in vitro* and *in vivo* systems, and (*ii*) characterize the early response pathways linked to toxicity in both rat *in vivo* and rat/human *in vitro* experiments. Building upon the recent study of Iskar et al. (Iskar et al. 2013) showing that

drugs affected modules of co-expressed genes conserved across a small set of three human cancer cell lines and rat liver samples, we developed a new pathway-based approach that combined gene set enrichment analysis (GSEA) and biclustering to efficiently integrate large-scale toxicogenomic data across different species. Our analysis showed that chemicals affect a set of conserved pathways linked to chemical-induced toxicity across species and experimental platforms.

MATERIAL AND METHODS

The overall design of our analysis is represented in Figure 1. The three experimental settings that we investigated in TG-GATEs are rat liver *in vivo*, rat and human primary hepatocyte *in vitro* and are referred to as RLV, PRH and PHH, respectively.

Microarrays retrieval and preparation

Rat liver, primary rat and human hepatocytes microarray data files were downloaded from ArrayExpress. The three studies with the accessions E-MTAB-799, E-MTAB-798, and E-MTAB-797 contain toxicogenomic data for rat liver *in vivo* (RLV), cultured primary human hepatocytes (PRH), cultured primary human hepatocytes (PHH) and experiments, respectively, for more than 100 chemical compounds (Figure 1A). PHH and PRH were treated with each compound in duplicates, using three increasing doses (low, middle and high doses) for three different amounts of time (2, 8 and 24 hours; Figure 1A). Rat liver samples were obtained from animals treated with each compound in triplicates and sacrificed at 3, 6, 9 and 24 hours after dosing (Figure 1A). The highest dose refers to the maximally tolerated dose. Each compound is associated with a corresponding vehicle control for all experimental conditions.

All CEL files (Affymetrix data format that contains the raw intensity values for both perfect match and mismatch probes) were checked for duplicated names and inconsistencies. For 71 chemicals, it was noted that the data from human hepatocytes treated with a low dose of compound was missing; these 71 chemicals were nevertheless retained and analyzed with the other 48 chemicals. In total, the transcriptional effects of 119 chemicals on human hepatocytes were gathered from 2,004 microarrays (Affymetrix GeneChip Human Genome U133 Plus 2.0 platform). Similarly, the effects of 129 chemicals on rat liver samples and rat hepatocytes were deduced from 6,192 and 3,096 microarrays, respectively (Affymetrix GeneChip Rat Genome 230_2.0) (Figure 1B). All datasets, including kidney samples in E-MTAB-799 and the repeated dose study (accession E-MTAB-800), are downloaded and curated on the fly through our fully automated pipeline. Documented code is available on GitHub (https://github.com/bhklab/TGGATES).

Gene expression data

Gene expression data were normalized with the robust multi-array average algorithm (RMA) (Irizarry et al. 2003) using the Bioconductor package BufferedMatrixMethods (version 1.30.0) (Gautier et al. 2004). Probes were mapped to Entrez Gene IDs using the Bioconductor annotation packages hgu133plus2.db (version 3.0.0) and rat2302.db (version 3.0.0) for human and rat, respectively. In case of multiple probes mapped to the same Entrez Gene ID, we used the Bioconductor package genefu (version 1.15.0) to select the most variant probeset for each gene. This procedure yielded 20,590 and 14,462 unique genes for human and rat, respectively.

Pathway collections

Every gene in the curated microarray experiments in TG-GATEs was assigned to pathways described in the Reactome database (Croft et al. 2014) using the Bioconductor package BiomaRt (version 2.22.0), for both rat and human genes present in the microarray platform. Pathway collection was performed on March 5, 2014. We subsequently selected the common pathways for rat and human, and retained only gene sets of sizes between 15 and 500 genes, which resulted in 419 common Reactome pathways for the GSEA analysis (Supplemental Material, Figure S1). For reproducibility, all curated pathways were stored in gmt files provided in https://github.com/bhklab/TGGATES.

Gene-chemical associations

Gene ranking was based on gene-chemical associations, which were identified by fitting linear models to estimate the effect of chemical dosage on gene expression levels controlled by treatment time and interaction between dosage and time. For each pair of gene *i* and chemical *j*, we used the following model

$$G_i = \beta_0 + \beta_1 D_j + \beta_2 T_j + \beta_3 D_j T_j \qquad \text{(Equation 1)}$$

where G_i denotes the expression value of gene *i*, D_j is the dose of chemical *j*, T_j is the treatment time with chemical *j*, β_0 is the intercept, β_1, β_2, and β_3 are the regression coefficients for the chemical dosage, treatment time, and interaction term of dose and treatment, respectively. The strength of the linear gene-chemical association is given by β_1 and its significance (*p*) is computed using Student *t* test as provided by the lm() function in the R stats package (Team 2013).

Pathway-chemical associations

Pathways that are significantly perturbed by each chemical were identified using the java implementation of GSEA (Subramanian et al. 2005) (version 2.0.14) provided by the Broad Institute. For each chemical, we first ranked all genes with respect to the signed significance of their gene-chemical association, that is sign(β_1) * $-\log_{10}(p)$ as in (Equation 1). We then used each chemical-specific ranked list of genes to perform a pre-ranked GSEA to calculate normalized enrichment scores (NES) for all common pathways between human and rat. The higher the absolute value of NES, the more enriched is the corresponding pathway in genes whose expression is significantly perturbed by the chemical of interest. We repeated this process for each chemical and created an "enrichment matrix" with pathway enrichment scores in rows and chemicals in column for each dataset (Figure 1B).

Conserved transcriptional modules

One hundred and fifteen chemical compounds were common to all three experimental settings (Figure 1B and Supplemental Material, List of Common Chemicals). For each of these datasets, we applied a biclustering method, that is the iterative signature algorithm (ISA) (Bergmann et al. 2003) implemented in the R package isa2 (version 0.3.3) (Csardi et al. 2010), on the enrichment matrix to simultaneously identify similar biochemical-induced transcriptional response patterns. The ISA algorithm runs with all combinations of threshold values on rows and columns, as described in details on the companion website (https://www.pmgenomics.ca/bhklab/pubs/tggates). Similarly to (Iskar et al. 2013) we merged modules with similar set of pathways using function *isa.unique()* in the *isa2* package to filter redundant modules using a correlation limit of 0.5 to determine redundant biclusters. Lastly, modules sharing common sets of pathways and chemical across the different datasets -- namely

RLV, PRH, and PHH (inter-dataset similarity) -- were identified using a one-sided hypergeometric test ($p < 1E-3$); this technique is referred to as the reciprocal best-hit approach (Iskar et al. 2013).

Reproducible research

To ensure full reproducibility, this work complies with the guidelines proposed by Robert Gentleman (Gentleman 2005) in terms of availability of the code and reproducibility of results and figures. The procedure to properly set up the software environment and run our analysis pipeline is provided in Supplemental Material, Reproducibility of analysis. The analysis code is also publicly available on https://github.com/bhklab/TGGATES.

RESULTS

The approach we used to investigate the pathways altered by chemical perturbations leverages the transcriptional profiling data available in TG-GATEs for rat liver *in vivo* (RLV) and for rat and human primary hepatocytes cultured *in vitro* (PRH and PHH, respectively), as summarized in Figure 1A. We analyzed each of these three datasets separately and compared the results from the *in vitro* treated hepatocytes (PRH and PHH) to those from the liver of treated rats (RLV), as this animal model is considered the gold standard in toxicity studies. Pre-processing of these gene expression datasets yielded a set of 20,590 and 14,460 unique genes from the human and rat microarray platforms, respectively, that were kept for subsequent analysis. The association between gene expression levels and the 115 chemicals common across the three TG-GATEs experimental settings (PRH, PHH, and RLV) was then investigated at the pathway-level using the pre-ranked version of gene set enrichment analysis (GSEA) (Subramanian et al. 2005). This was done with 419 pathways, which were in common between rat and human

organisms as queried from Reactome database, in order to identify modulated pathways upon chemical perturbation (Supplemental Material, Figure S1). Matrices containing the enrichment scores of each pathway perturbed by each chemical were then analyzed using an unsupervised biclustering technique called Iterative Signature Algorithm (ISA) (Bergmann et al. 2003) to define functional modules (i.e., clusters of pathways) that are specifically associated with diverse chemical treatments. Each module is given a summary name, that is a Reactome parent term that best recapitulates the pathways enriched in this module (see Supplemental Material, Table 1).

Conservation of transcriptional modules across experimental settings

Rat liver in vivo (RLV) treated with a single dose: Twenty-four non-redundant modules were identified using the aforementioned ISA analysis ($p < 1E-3$). These modules were enriched for the following biological pathways: neuronal system, hemostasis, cell cycle checkpoints, DNA repair, mitosis, lysosomes disorders, innate immune system, NOTCH, TGF-βR/SMAD and PI3K/AKT signalling cascades, lipid metabolism, and mitochondrion dependant processes. The summary names of all modules are provided in Supplemental Material, Table 1.

Primary Rat hepatocytes (PRH) vs. Rat Liver in Vivo (RLV): The ISA algorithm detected eighteen modules in PRH. Interestingly, seventeen modules overlapped with RLV using a reciprocal best-hit approach in which two modules are considered as conserved if their Reactome pathways significantly overlap (Iskar et al. 2013) (hypergeometric, $p <1E-3$). Only one module related to cholesterol biosynthesis did not overlap at the considered cutoff. Figure 2 shows in detail the number of non-redundant ISA modules in each dataset and the conservation across the experimental settings.

Primary human hepatocytes (PHH) vs. Rat Liver in Vivo (RLV): ISA analysis resulted in the identification of fifteen modules in PHH toxicogenomic data. Again, fourteen of them overlapped with RLV (hypergeometric, p <1E-3; Figure 2).

Overall, we identified thirteen modules to be conserved across the three experimental setting datasets (RLV, PHH, and PRH). (Supplemental Material, Table 1 ; Figure 2). As a representative example, we show a conserved module in Figure 3. It is enriched for components of the innate immune system, with the overlapping pathways in the same order for both RLV, PHH and PRH. We extracted the union of the genes that were found to contribute to the enrichment score (referred to as *leading edge;* (Subramanian et al. 2005)) of at least one pathway for all chemicals in the module. From this, we obtained a list of common genes that are activated or repressed by chemical stress in RLV, PRH and PHH (Heatmaps for all ISA settings, lists of hypergeometric p-values, and lists of leading edge genes are provided in separate Supplemental Material files S2, S3, S4).

Enrichment for hepatocarcinogens

The approach described above identified thirteen modules that are associated with the early response of hepatocytes to diverse chemicals and are conserved *in vivo*, *in vitro* and between rat and humans. To test if some modules were significantly associated with the hepatocyte response to known hepatocarcinogens, we investigated twenty five previously validated rat hepatocarcinogens (Yamada et al. 2012) present among the 115 chemicals investigated in our study (Supplemental Material, List of common chemicals). Specifically, these hepatocarcinogens were significantly enriched in the NOTCH and TGF-βR/SMAD signaling modules in PHH (hypergeometric $p < 0.05$), but not in PRH or RLV. The TGF-βR/SMAD

signaling module (Figure 4A) in PHH was enriched for known environmental toxicants and carcinogens (e.g., ethionine, thioacetamide, coumarin, ethanol, acetamidofluorene, nitrosodiethylamine). None of these modules was enriched for hepatocarcinogens in RLV and this was only the case for the PI3K/AKT (Phosphoinositide 3-kinase) module in PRH (p = 0.049; see Supplemental Material). The known rat hepatocarcinogens were also significantly associated with the neuronal system/G protein-coupled receptors (GPCRs) module in both RLV and PHH, but not PRH, probably reflecting the pleiotropic roles that GPCRs play in many cellular processes, including chemical carcinogenesis (see Supplemental Material S2 and S3).

As a control experiment, we selected twelve non-carcinogenic compounds (Supplemental Material, List of common chemicals), and determined if they were significantly associated with any of the modules in RLV, PRH and PHH. As anticipated, no enrichment was observed, especially for those modules enriched for known hepatocarcinogens in PHH. As an additional control, we ascertained that the NOTCH and TGF-βR/SMAD modules were indeed enriched in cancer-related pathways; this was done by showing that the 20 pathways containing the word "Cancer" in the Reactome common dataset (over a total of 419 pathways) were in fact enriched in those modules (hypergeometric, $p < 1E-3$). This was not the case for any of the remaining modules without cancer terms. Collectively, the results presented above support that primary human hepatocytes can detect potential environmental chemical carcinogens (Figure 4A). By extension, we infer that the other modules are also enriched in pathways that are pertinent to chemical exposure.

Activation of the Peroxisome proliferator activated-receptor alpha (PPARalpha)

Since some PPARalpha activators are known to induce hepatocarcinogenesis in rodents' liver, we tested if PPARalpha activators (e.g. benziodarone, benzbromarone, fenofibrate, clofibrate,

ibuprofen, WY14643, and gemfibrozil) were randomly distributed across modules in RLV, PHH and PRH. Interestingly, none of the modules in PHH or PRH were enriched for those drugs, however we found that a module unique to RLV was significantly associated with the regulation of lipid metabolism by PPARalpha and enriched for those drugs ($p = 0.014$). Other PPARalpha potential inducers were found in this module including non-steroidal anti-inflammatory (NSAIDs) and anti-tuberculosis drugs (Figure 4B).

A recent study (Grinberg et al. 2014) showed that numerous compounds from TG-GATEs cause ”stereotypical” transcriptional responses in PHH. Such definition is given when a cytotoxic concentration of numerous compounds caused a consensus expression response regardless of the chemical class of compound. We assessed the significance of the overlap, for each module, between all leading edge genes, which we generated from the biclustering in PHH, and the deregulated genes by at least 20 compounds in their study. We demonstrated that “stereotypical” clusters of genes, involved in liver metabolic functions and cell proliferation, were enriched in two modules from PHH, mainly those associated with normal liver function and DNA synthesis modules. Furthermore, to ascertain that our observations from PHH are not simply experimental artefacts due to *in vitro* conditions, we selected liver cirrhosis as a case study and tested the enrichment for genes associated exclusively with liver cirrhosis in PHH (Grinberg et al. 2014). Interestingly, the transforming growth factor beta-receptor signaling module in PHH (TGF-βR, module 6) was significantly enriched for genes linked to liver cirrhosis besides being induced by known hepatocarcinogens and environmental toxicants (Figure 4B).

Finally, we showed that the distribution of genes perturbed by rat hepatocarcinogens vs. non-hepatocarcinogens was alike (see Supplemental Material, Figure S2).

DISCUSSION

We tested the extent to which transcriptional responses associated with liver toxicity can be recapitulated across human and rat and between *in vivo* and *in vitro* settings. To do so, we exploited the toxicogenomic information generated by the TG-GATEs project, from liver samples of rats treated with different chemicals and from rat/human hepatocytes exposed to the same compounds *in vitro*. To date, several studies have used TG-GATEs to build predictors of relevant toxicological endpoints. For example, Zhang et al. recently used this data to build a predictive gene signature for both hepatotoxicity and nephrotoxicity (Zhang et al. 2014). Interestingly, this study revealed the importance of early response genes in triggering toxicity-associated signalling networks, as highlighted by the high predictive power of the signature derived from a treatment period of less than 24 hours.

To our knowledge, our study is the first analysis of the TG-GATEs data comparing the functional changes - in the form of transcriptional responses - that are induced by a large panel of chemicals *in vivo* (rat liver)*, in vitro* (cultured hepatocytes) and between species (human vs. rat). A main feature of our approach is the fact that it relies on a pathway enrichment analysis, thus allowing comparison to be made between species without having to rely on the limited subset of orthologous genes. In this context, it is worth contrasting our findings to those of Iskar et al. (Iskar et al. 2013), who identified, solely based on a orthologous genes, transcriptional modules that were conserved between rat liver (Natsoulis et al. 2008) and three human cancer cell lines from the Connectivity Map (CMap) (Lamb et al. 2006). Their findings showed that 15% of the chemical-induced modules were conserved across cell lines and species. However, this approach was limited to 8,962 genes in CMap, which corresponded to only 3,618 orthologous genes available for the rat liver experiments. To overcome this

limitation, by focusing on common pathways between species, our approach enabled a full exploration of the TG-GATEs datasets and the identification of functional pathways altered by chemical treatments in both rat and human.

Our results indicate that the response of hepatocytes to chemical insults is analogous *in vitro*, *in vivo*, and across human and rat in that it involves a conserved set of cellular pathways. Specifically, we identified thirteen highly conserved modules representative of the early response of hepatocytes to chemical exposure. Two of those are enriched in key signalling pathways associated with cancer, namely the transforming growth factor beta receptor superfamily module (TGF-βR -mod17 RLV) the NOTCH signaling module (NOTCH-mod6 RLV). Given the role that the TGF-βR and NOTCH pathways play in response to early toxicity (Zhang et al. 2014) and in maintaining normal liver functions (Morell and Strazzabosco 2014), respectively, it was not surprising that these modules were enriched for known rat hepatocarcinogens including environmental toxicants. What could be more puzzling, according to our results, is the fact that these two pathways are significantly associated with hepatocarcinogens only in human and not in rat. This may reflect a key difference in how both species deal with these chemicals. That the response of rats and humans may differ for some chemicals is also supported by our finding that the PPARalpha agonists clofibrate, fenofibrate, gemfibrozil, benziodarone, and benzbromarone up-regulate pathways associated with PPARalpha activation only in rat liver, thus providing a potential mechanism underlying the hepatocarcinogenicity of these drugs in rats but not humans (Lai 2004).

Several lines of evidence suggest that the modules identified in this study are relevant to how hepatocytes respond to chemicals. For example, one of the modules we identified, the innate immune system (mod2 RLV), was enriched in pro-inflammatory Toll-like receptor signalling

pathways, which have been shown by Huang et al. to be good predictors of drug induced liver injury (Huang et al. 2010). Our results are also consistent with those reported in the comparative studies of Doktorova et al., who assessed the transcriptional profiles of toxicants between rat liver and a panel of *in vitro* models (Doktorova et al. 2012; Doktorova et al. 2013). Those studies assigned deregulated genes from *in vivo/in vitro* comparison. Moreover, we found that pathways associated with G protein-coupled receptors (GPCRs) and the neuronal system were consistently affected by a variety of chemicals. Of particular relevance is the fact that some chemicals found in this conserved module (neuronal system-mod8 RLV) can cause the potentially lethal long QT syndrome (delayed repolarization of the heart) by perturbing heart conductance. For example, ciprofloxacin, haloperidol, thioridazine, quinidine, and amiodarone are well known to prolong the QT interval and cause *Torsades de Pointes,* a deadly form of arrhythmia (Fazio et al. 2013). This module was also enriched for known rat hepatocarcinogens in RLV and PHH but not PRH, a finding that may relate to the fact that ion channels, in addition to being involved in the long QT syndrome, can also play a role in carcinogenesis (Babcock and Li 2013). However, this observation might not be specific to a class of compounds since the Reactome pathways related to the neuronal system contain a large number of genes (> 500). Our findings also suggest that some chemicals modulate pathways associated with vitamin metabolism (metabolism of vitamins and cofactors-mod3 in RLV) in hepatocytes, in particular those associated with the inherited metabolic disorders ethylmalonic aciduria and homocystinuria. Surprisingly, the scientific literature contains only a few reports pertaining to the association between chemical-induced liver injury and vitamins. Amongst the studies that we found relevant to this work, one describes an association between high levels of circulating cobalamin (vitamin B12) and several serious liver diseases (Ermens et al. 2003),

while the other highlights the role of vitamin B12 metabolism in Methylmalonic aciduria, a disorder that can lead to severe liver injury and require in some cases a liver transplantation (Hansen and Horslen 2008). Given the strong association between vitamin metabolism and early drug exposure revealed in our study, it may be of interest to explore further this understudied area of research.

Furthermore, we further confirmed the biological relevance of our biclusters against a recent study (Grinberg et al. 2014). Indeed, we showed that our modules recapitulated stereotypic response to chemicals as well as compound-specific perturbations. Moreover, we found evidence that the TGF-β receptor signalling module in PHH could act as a potential biomarker of chemical injury that may lead to liver cirrhosis besides being enriched for known hepatocarcinogens.

It is worth mentioning that our new bioinformatics pipeline complements previous approaches, used to elucidate the mechanisms of chemical toxicity *in vitro* or *in vivo*, by enabling efficient and unbiased exploration of chemical-induced transcriptional changes in both *in vivo* and *in vitro* systems, and across different species. The modules that emerged from this analysis suggest that functional networks of xenobiotic detoxification and response to external stress are highly conserved in the hepatic system across human and rat. In contrast to pathway conservation, our results suggest that the chemicals associated with any given module, do not show a meaningful overlap between *in vitro* and *in vivo* systems or across species. Although somewhat counter-intuitive, this has been observed previously (Zhang et al. 2014) and may reflect *bona fide* differences in chemical bioactivation through metabolism between systems, thus complicating the interpretation of *in vivo* versus *in vitro* data. Another factor to consider when assessing the value of our approach is the fact that it relied on an expert knowledge

curated, peer-reviewed database of functional pathways. While it provided an alternative resolution for the orthologous gene limitation, we are nevertheless aware that annotations in pathway databases are incomplete and thus may limit this approach to some extent. Some of these limitations may be addressed in the future as we extend our approach to other systems (e.g., HepG2 hepatocellular carcinoma cell line), other toxicogenomic databases, such as DrugMatrix (Natsoulis et al. 2008), and integrate more 'omics' data including RNA-seq, and SNP profiling to take into account the variability of individual response to chemicals.

CONCLUSION

The analysis of the TG-GATEs data presented here indicates that toxicogenomics-based cellular models recapitulate most of the pathways related to chemical-induced injury in rat liver. Furthermore, it may be possible to reduce unnecessary animal testing in early toxicological assessments and complement them with *in vitro* testing. Because environmental toxicants can be associated with alterations in cellular pathways that contribute to general injury patterns and likely more severe phenotypes including carcinogenesis, we showed that the TGF-βR/SMAD module could serve as a putative biomarker to identify chemicals with carcinogenic potential for humans. Especially that potent carcinogenic compounds such as 2-acetamidofluorene, nitrosodiethylamine and ethanol were found in this module in PHH.

Our findings could be generalized to study a large set of environmental contaminants relevant to human health. Therefore, our method helps identify numerous pathways and genes that are responsible of toxicity controlled by chemical exposures.

REFERENCES

Abbott A. 2005. Animal testing: More than a cosmetic change. Nature 438:144-146.

Afshari Ca, Hamadeh HK, Bushel PR. 2011. The evolution of bioinformatics in toxicology: Advancing toxicogenomics. Toxicological sciences : an official journal of the Society of Toxicology 120 Suppl:S225-237.

Babcock JJ, Li M. 2013. Herg channel function: Beyond long qt. Acta pharmacologica Sinica 34:329-335.

Bergmann S, Ihmels J, Barkai N. 2003. Iterative signature algorithm for the analysis of large-scale gene expression data. Physical review E, Statistical, nonlinear, and soft matter physics 67:031902.

Bissell DM, Gores GJ, Laskin DL, Hoofnagle JH. 2001. Drug-induced liver injury: Mechanisms and test systems. Hepatology 33:1009-1013.

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. 2003. Arrayexpress--a public repository for microarray gene expression data at the ebi. Nucleic acids research 31:68-71.

Caiment F, Tsamou M, Jennen D, Kleinjans J. 2014. Assessing compound carcinogenicity in vitro using connectivity mapping. Carcinogenesis 35:201-207.

Chen M, Zhang M, Borlak J, Tong W. 2012. A decade of toxicogenomic research and its contribution to toxicological science. Toxicol Sci 130:217-228.

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. 2014. The reactome pathway knowledgebase. Nucleic acids research 42:D472-477.

Csardi G, Kutalik Z, Bergmann S. 2010. Modular analysis of gene expression data with r. Bioinformatics 26:1376-1377.

Doktorova TY, Ellinger-Ziegelbauer H, Vinken M, Vanhaecke T, van Delft J, Kleinjans J, et al. 2012. Comparison of genotoxicant-modified transcriptomic responses in conventional and epigenetically stabilized primary rat hepatocytes with in vivo rat liver data. Arch Toxicol 86:1703-1715.

Doktorova TY, Yildirimman R, Vinken M, Vilardell M, Vanhaecke T, Gmuender H, et al. 2013. Transcriptomic responses generated by hepatocarcinogens in a battery of liver-based in vitro models. Carcinogenesis 34:1393-1402.

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ. 2008. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutation research 637:23-39.

Ermens AA, Vlasveld LT, Lindemans J. 2003. Significance of elevated cobalamin (vitamin b12) levels in blood. Clinical biochemistry 36:585-590.

Fazio G, Vernuccio F, Grutta G, Re GL. 2013. Drugs to be avoided in patients with long qt syndrome: Focus on the anaesthesiological management. World journal of cardiology 5:87-93.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. Affy--analysis of affymetrix genechip data at the probe level. Bioinformatics 20:307-315.

Gentleman R. 2005. Reproducible research: A bioinformatics case study. Statistical applications in genetics and molecular biology 4:Article2.

Greaves P, Williams A, Eve M. 2004. First dose of potential new medicines to humans: How animals help. Nature reviews Drug discovery 3:226-236.

Grinberg M, Stober RM, Edlund K, Rempel E, Godoy P, Reif R, et al. 2014. Toxicogenomics directory of chemically exposed human hepatocytes. Arch Toxicol 88:2261-2287.

Hansen K, Horslen S. 2008. Metabolic liver disease in children. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society 14:713-733.

Hebels DG, Jetten MJ, Aerts HJ, Herwig R, Theunissen DH, Gaj S, et al. 2014. Evaluation of database-derived pathway development for enabling biomarker discovery for hepatotoxicity. Biomarkers in medicine 8:185-200.

Hirode M. 2008. Gene expression profiling in rat liver treated with various hepatotoxic-compounds inducing coagulopathy. 2002.

Hirode M, Ono A, Miyagishima T, Nagao T, Ohno Y, Urushidani T. 2008. Gene expression profiling in rat liver treated with compounds inducing phospholipidosis. Toxicology and applied pharmacology 229:290-299.

Huang J, Shi W, Zhang J, Chou JW, Paules RS, Gerrish K, et al. 2010. Genomic indicators in the blood predict drug-induced liver injury. Pharmacogenomics J 10:267-277.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249-264.

Iskar M, Zeller G, Blattmann P, Campillos M, Kuhn M, Kaminska KH, et al. 2013. Characterization of drug-induced transcriptional modules: Towards drug repositioning and functional understanding. Mol Syst Biol 9:662.

Kier LD, Neft R, Tang L, Suizu R, Cook T, Onsurez K, et al. 2004. Applications of microarrays with toxicologically relevant genes (tox genes) for the evaluation of chemical toxicants in sprague dawley rats in vivo and human hepatocytes in vitro. Mutat Res 549:101-113.

Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? Nature reviews Drug discovery 3:711-715.

Lai DY. 2004. Rodent carcinogenicity of peroxisome proliferators and issues on human relevance. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 22:37-55.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313:1929-1935.

Metushi IG, Uetrecht J. 2014. Lack of liver injury in wistar rats treated with the combination of isoniazid and rifampicin. Molecular and cellular biochemistry 387:9-17.

Morell CM, Strazzabosco M. 2014. Notch signaling and new therapeutic options in liver disease. Journal of hepatology 60:885-890.

Natsoulis G, Pearson CI, Gollub J, B PE, Ferng J, Nair R, et al. 2008. The liver pharmacological and xenobiotic gene response repertoire. Mol Syst Biol 4:175.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. 1999. Microarrays and toxicology: The advent of toxicogenomics. Molecular carcinogenesis 24:153-159.

REACH. Http://www.Reachonline.Eu/reach/en/reach_en/preamble1.Html.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102:15545-15550.

Suter L, Schroeder S, Meyer K, Gautier JC, Amberg A, Wendt M, et al. 2011. Eu framework 6 project: Predictive toxicology (predtox)--overview and outcome. Toxicol Appl Pharmacol 252:73-84.

Team RDC. 2013. R: A language and environment for statistical computing. Vienna, Austria:R Foundation for Statistical Computing,URL http://www.R-project.org.

Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, et al. 2010. The japanese toxicogenomics project: Application of toxicogenomics. Molecular nutrition & food research 54:218-227.

Uehara T, Minowa Y, Morikawa Y, Kondo C, Maruyama T, Kato I, et al. 2011. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. Toxicol Appl Pharmacol 255:297-306.

Yamada F, Sumida K, Uehara T, Morikawa Y, Yamada H, Urushidani T, et al. 2012. Toxicogenomics discrimination of potential hepatocarcinogenicity of non-genotoxic compounds in rat liver. Journal of applied toxicology : JAT 2012:1284-1293.

Zhang JD, Berntenis N, Roth A, Ebeling M. 2014. Data mining reveals a network of early-response genes as a consensus signature of drug-induced in vitro and in vivo toxicity. Pharmacogenomics J 14:208-216.

FIGURE LEGENDS

Figure 1: Analysis workflow for the TG-GATEs dataset. (**A**) Overview of the TG-GATEs experimental design. TG-GATEs includes rat liver *in vivo* (RLV), rat hepatocytes *in vitro* (PRH), and human hepatocytes *in vitro* (PHH) experiments to test transcriptional responses for more than 100 chemical compounds. Samples have been treated with chemical with three doses along with a control group, and gene expression was measured repeatedly within 24 hours as shown. (**B**) Pathway-based analysis pipeline. A comparative analysis of the three TG-GATEs experiments was conducted by investigating chemical-induced pathways in RLV, PRH and PHH. For each chemical, a linear regression model was fitted for every gene to assess the effect of chemicals on gene expression levels taking into account treatment period and dose level. Based on these association models, genes were ranked to perform a gene set enrichment analysis (GSEA) on common Reactome pathways. From the enrichment results transcriptional modules conserved across experimental settings (RLV, PRH, and PHH) were identified by bi-clustering.

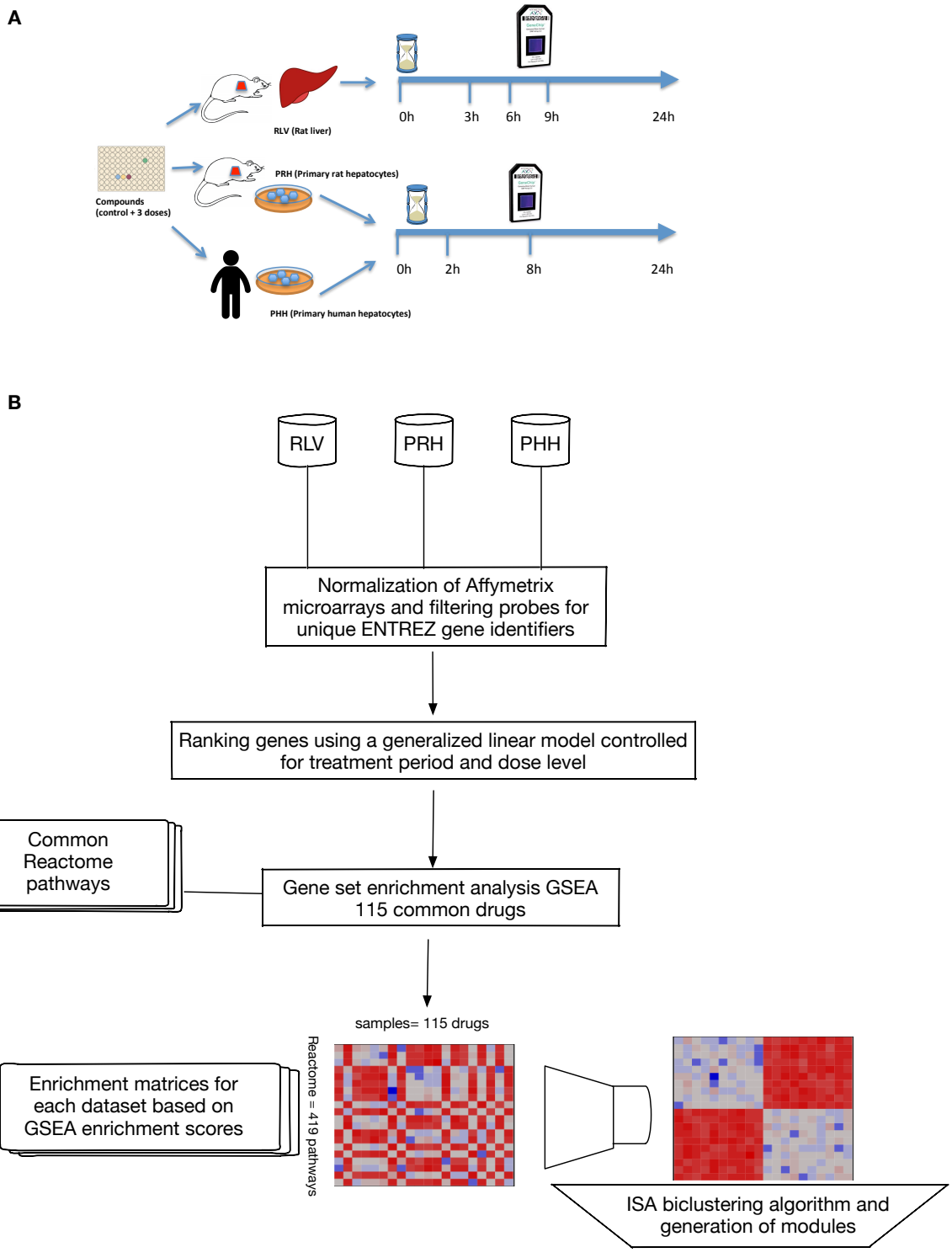
Figure 2: The number of non–redundant transcriptional modules and proportions identified for each and across all experimental settings in TG-GATEs. Each bar corresponds to an experimental setting in TG-GATEs (RLV, PRH, PHH) and contains the number of modules found to be unique for the experiment (blue) or show a corresponding module in another experiment (see color legend: Green for conserved, grey for conserved in RLV vs. PHH, orange for conserved in PHH vs. PRH, yellow for conserved in RLV vs. PRH). While only few modules were detected in only one or two settings, most modules showed significantly high overlap in terms of pathway enrichment between all settings (hypergeometric $p < 1E-3$).

Figure 3: Conservation of modules across *in vitro* and *in vivo* settings based on Reactome pathways. This example summarizes a conserved module between RLV, PRH and PHH, shown as heatmaps and keeping overlapping pathways colored with respect to their enrichment scores: up-regulated pathways are shown in blue and down-regulated are shown green. Three heatmaps corresponding to a conserved module associated with the innate immune system (mod2 in RLV, mod15 in PHH and mod10 in PRH). The leading edge genes from common

pathways, and activated or repressed by chemicals are shown under the heatmap with known oncogene colored in red. (More details in Supplemental Material S2 and S4).

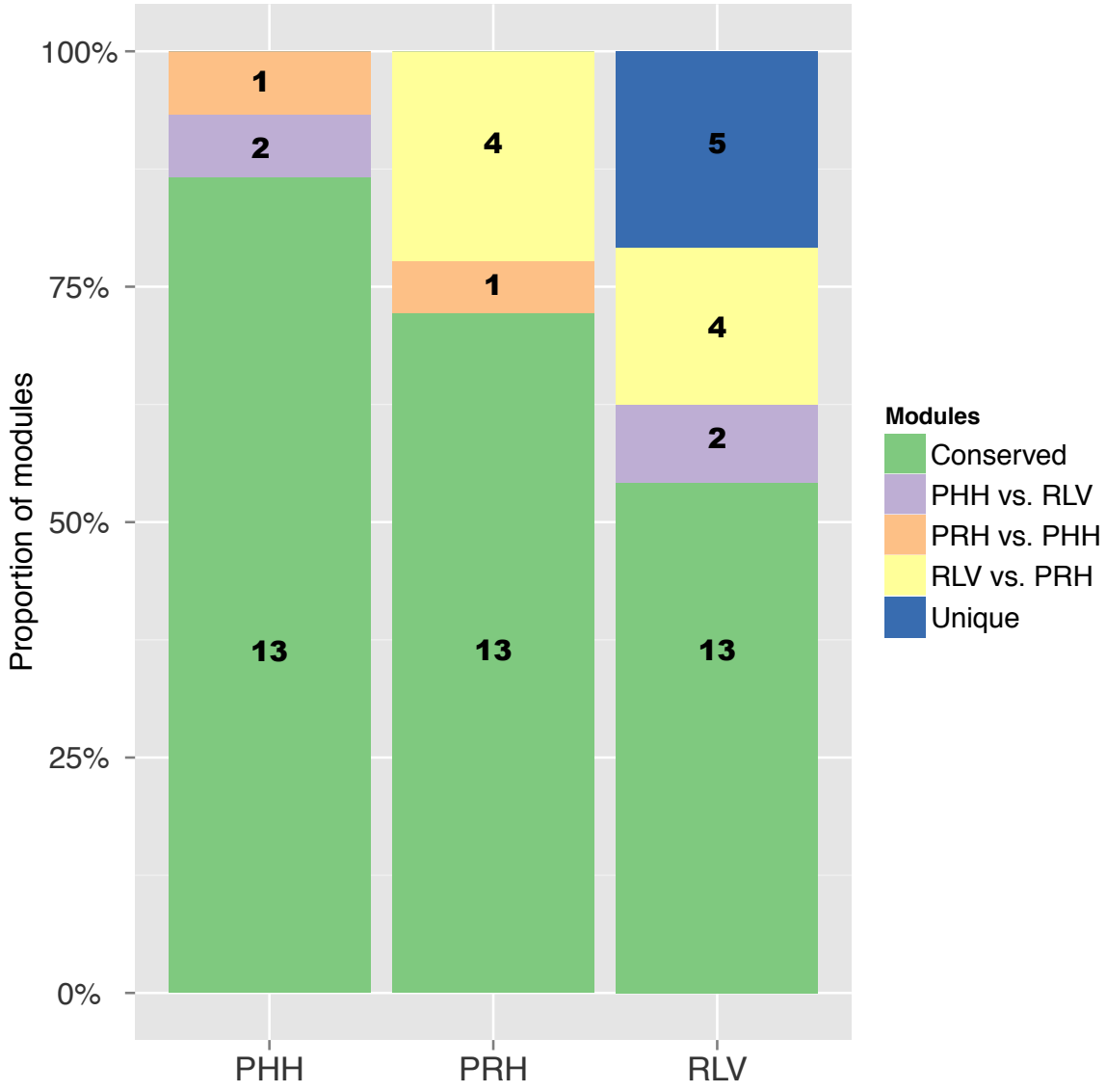
Figure 4: Characterization of putative biomarkers within chemical-induced modules. (**A**) Heatmap representing a module in PHH (mod6), associated with the transforming growth factor beta receptor signalling that can be considered as a candidate biomarker in humans for environmental exposure to known toxicants. Diverse rat hepatocarcinogens were enriched in this module. (**B**) Heatmap representing a module in RLV (mod5) that was relevant to toxicity mode of action. It is enriched for a class of lipid lowering drugs known as fibrates. These drugs are rat hepatocarcinogens and activate the peroxisome proliferation-activated receptor alpha (PPARA). Drugs that activate PPAR pathways include non-steroidal anti-inflammatory and anti-tuberculosis drugs. All statistical details and genes contributing to those pathways are found in Supplemental Material S4.

Figure 1.


A
RLV (Rat liver)
0h
3h
6h
9h
24h
Compounds
(control + 3 doses)
PRH (Primary rat hepatocytes)
0h
2h
8h
24h
PHH (Primary human hepatocytes)
B
RLV
PRH
PHH
Normalization of Affymetrix microarrays and filtering probes for unique ENTREZ gene identifiers
Ranking genes using a generalized linear model controlled for treatment period and dose level
Common Reactome pathways
Gene set enrichment analysis GSEA 115 common drugs
samples= 115 drugs
Enrichment matrices for each dataset based on GSEA enrichment scores
Reactome = 419 pathways
ISA biclustering algorithm and generation of modules

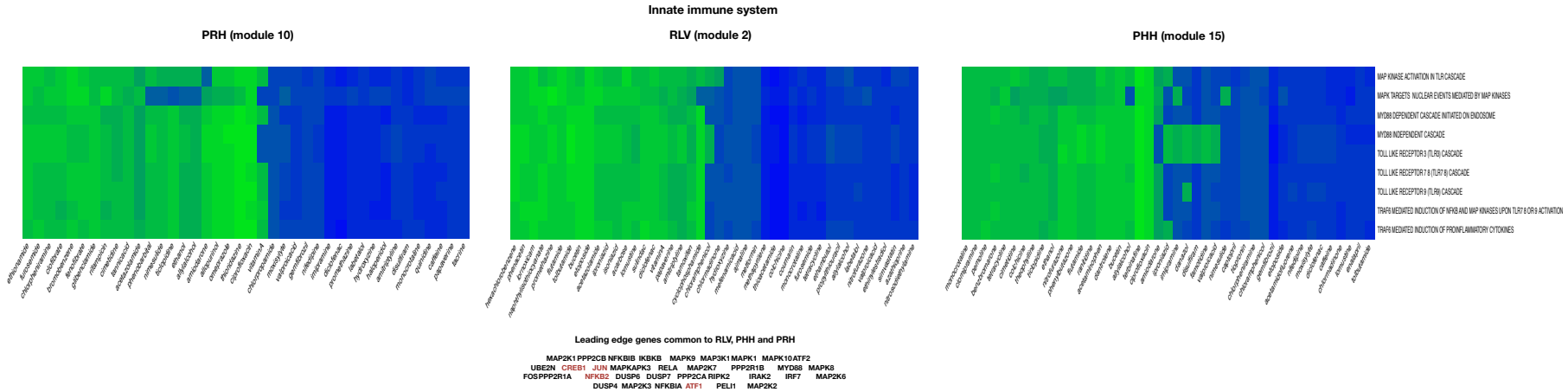
Environ Health Perspect DOI: 10.1289/ehp.1409157
Advance Publication: Not Copyedited

Figure 2.



Environ Health Perspect DOI: 10.1289/ehp.1409157
Advance Publication: Not Copyedited

Figure 3.



Environ Health Perspect DOI: 10.1289/ehp.1409157
Advance Publication: Not Copyedited

Figure 4.